DOCUMENT RESUME

ED 132 568 CS 203 088

AUTHOR Chisholm, William
TITLE Degrees of Syntactic and Rhetorical Fluency-Competency in Freshman Writing: A Computer-Assisted Study.
PUB DATE 77
NOTE 7p.; Paper presented at the Annual Meeting of the Midwest Modern Language Association (18th, St. Louis, Missouri, November 4-6, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS College Freshmen; *Composition Skills (Literary); Higher Education; Language Fluency; *Language Patterns; Language Research; *Rhetoric; *Syntax

ABSTRACT

An exploratory study of quantitative measurement of syntactic and rhetorical fluency examined students' writing near the beginning and near the end of a two-quarter, freshman English program. The syntactic analysis focused on the clause, which was classified according to basic syntactic type and elaborating syntactic structures. The rhetorical analysis concentrated on the orthographic unit and included counts of selected rhetorical features and counts of logical relationships between successive units of thought. Preliminary results are reported, though in general the measures chosen did not discriminate between the 20 compositions written at the beginning of the program and the 20 written at the end. (AA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED132568

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

William Chisholm

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

DEGREES OF SYNTACTIC AND RHETORICAL FLUENCY-COMPETENCY IN FRESHMAN WRITING: A COMPUTER-ASSISTED STUDY

William Chisholm
Cleveland State University

At the meeting of the Practical Linguistics Group here in St. Louis two years ago, I was pleased to outline the theory behind the Freshman English program at Cleveland State. That theory is based on the observation that a student's intuited linguistic knowledge can be made useful to him if he can comprehend it directly, if he can get it out of his left hemisphere and into the hand that holds his pen. The argument is simple: since a student already uses metaphor, logic, syntactic variety and embedding, and most of the other elements of linguistic structure, he needs to learn how to control these elements and to adapt them to the conventions of writing.

I said, too, that we were planning to investigate the syntactic and rhetorical structures that students write, and speculated that there might be a quantitative measure of "fluency" for syntax and rhetoric. I'm here today to report the results of the first phase of that investigation.

From the beginning, we understood the need to identify and describe the optimally fluent English sentence. Such a datum would allow us to describe the ways in which student sentences differ from the "average" professional sentence and provide the facts, perhaps, on which a goal-directed instruction could be based. By sampling a million sentences from representative prose writers, one could unearth, let's say, the "average sentence" of respected writing. One might begin this task by calculating the average number of words per sentence. Then one might figure out how many embeddings there are per sentence (relatives, infinitives, indirect questions, etc.) Next one could count the noun phrases, prepositional phrases and clauses. Over on the rhetoric side, though the problems are different and more difficult because the basic rhetorical unit is sometimes larger, sometimes smaller than the orthographic unit, one could count sentence connectors, inversions, analogies, conscious metaphors, latinate words and the rest, and get an average number of these per rhetorical unit. So then we would have some numbers: for syntax, say, 13 words per orthographic unit, 1.3 clauses, 1.0 embeddings per clause; and for rhetoric, .25 sentence connectors, .10 inversions, .05 analogies, .05 conscious metaphors, 22.0 % latinate forms...

Unfortunately, the exercise would give statistical anthropological data whose explanatory power was feeble. Why? For one thing, no sentence of English could be concocted that had 1.2 clauses. Still, we could determine to what extents the average freshman sentence deviates from the average professional sentence. For another, the standard average English sentence (SAES) wouldn't tell us much about why one author writes effectively but another does not. Effective writing depends importantly on variety: more than two or three sentences in a row with thirteen words in each makes for bad writing. Besides, syntactic or rhetorical complexity leads to poor writing as often as to good writing. Rhetorical complexity itself sometimes weakens the prose.

CS203088

At this point, it would be tempting to expand the analysis. For instance, one might attempt to find out what sentence-length variety is. Knowing this, we wouldn't be constrained merely to observe that student sentences deviate from standard average length. We might write another program (because our imaging was sloppy) that defines average sentence length as anything from 9-17 (4 on either side of 13), short as anything from 1-8, and long as anything above 17--and ask the computer to array sentence length in three ways and to calculate the intervals. We could continue to elaborate the analysis, down to getting coefficients for the relationship between the incidence of relative clauses, say, and logical connectors. We could dissect to the bone and bare the marrow, but would such calculations tell us what a good sentence is, or paragraph? No. We would end up as the political analyst does, knowing everything there is to know about the voter except how he or she will vote. Nevertheless, there are some temptations to go about the investigation in this manner, and some justification. If we remind ourselves that our first need is find out what the SAES is, not the "good" English sentence, then the task needs no other justification. The proper questions are: How many words are there in the SAES, how many clauses, how many embeddings?

Let me just mention here what I consider to be the most troubling pedagogical problem. Suppose in a thoroughgoing analysis of the syntax and rhetoric of renowned writers and of freshmen we found such comparative facts as the ones displayed in the chart following.

Structures	Professional Writing	Freshman Writing
Average number of sentences per paragraph	10	5
Average number of words per paragraph	200	50
Semantic field of nouns	narrow	wide
Correlation of numbers of persuasive sentences and illative relations	high	low
Correlation between numbers of persuasive/illative sentences and additive/illustrative sentences.....	high	low

What might we do with such relationships, pedagogically? How might we interpret them? All we might do with the facts is fit them into some available theory of the teaching of composition skills, or theorize anew. (The question about interpretation is subsequent, I think, to further fact gathering.) The previous question is the one that has interested me, namely, what is the syntactic and rhetorical structure of a sample of freshman writing? Despite the objections that I have just raised, it was this question that interested us. I wish to relate how we went about answering it.

We analyzed a sample of freshman writing to determine what the quantities were. And we have set the analysis of professional writing aside. The sample

itself consisted of twenty impromptu compositions written by entering frehsmen during the first week of instruction, and twenty more (different students) written toward the end of our two quarter program. No thoroughgoing effort was made to get a valid sample, though we did get themes that were written by as many women as men under roughly the same circumstances. Every fifth one from different teachers' sets of compositions was chosen until ten written by men and ten by women had been selected for each of the two sets.. Illegible themes after Xeroxing were thrown out and the next in the batch (the sixth) was put into the set.

In this way, by choosing as many compositions from one teacher's class as from another, and as many from an eight o'clock class as from a ten o'clock, we thought to reduce contamination of the sample. No dependent variables were controlled at all: we do not know how well the students did in the courses, how many books they read, how much T.V. they watched, or anything else like this.

The analysis was conducted in the following way. First we isolated the basic syntactic unit, the clause. The unelaborated minimal clause ranged from one word imperatives to N V N N structures like "The turtle gave the twig a tug." Each clause in each sentence was recorded by basic syntactic type--V, NV, N LV N, and so forth. We listed the elaborating syntactic structures that we would record: adverb, prepositional phrase, relative clause, noun phrase complement, infinitive phrase, etc. (Appendix A) Weights were assigned these special features in an ad hoc fashion, but we felt safe in assigning higher scores to double-based transformations than single, higher scores to adverbial phrases than to single-word adverbs, etc. We trained student assistants to identify these structures and got them to analyze the compositions and report such facts as these:

```
Composition   #1
Clause    #6
Words   16
Type   N LV ADJ
Elaborations:  adverb, predicate adjective complement, relative clause,
               triple noun...
```

This information was punched on cards and stored on tape.

For the rhetoric, the basic structure was the orthographic unit. We counted the following rhetorical features (as we found them): parallel structure ("...people were stopped in the street and approached in cafes; asyndeton ("...he was a friend, a lover, a mirage..."); inversion, but only subjects, verbs, objects, and certain prepositional phrases ("...algebra I know nothing about..."); appositives ("...his uncle, a taxidermist, lives in Maine..."); parenthesis ("...he drives his Volvo (Why am I telling you this?), insouciantly ..."; rhetorical ellipsis (...she was bursting with ideas, but timid"; repetition ("... she smiled at his half-smile, smiled at all he did..."; anaphora, conscious metaphor, analogy,

and antithesis .* Also, we have marked the logical relationships between successive units of thought: additive, adversative, illative, causative, illustrative, alternative, causal or nil.

Then the computer. Using SPITBOL (Version 2.2) we wrote an arithmetic program that could report averages per composition and set for the items counted (by clause, by orthographical unit, and by paragraph.) We wanted to know, for instance, what the average number of embeddings per clause was, what the average number of parallel structures per paragraph was, etc. Computers are good with this sort of thing. We got our numbers (Appendix B).

Computers are better at a related activity--statistical evaluation. We used two different evaluation procedures. We submitted our numbers first to the SPSS (Statistical Package for the Social Sciences), and discovered that some correlations were nearly perfect: when we applied T-tests to the means and variances that the SPSS provided we proved that the raw statistical differences were not significant. There's only one chance in a thousand, for example, that the number of embeddings per clause will not go up as the number of words per clause goes up. Similarly, we would have got 0.91 embeddings for one set and 1.15 for the other (the actual figures) 50% of the time, that is, just by sampling error. In this case as in similar statistical cases, 0.91 equals 1.15.

Pearson Correlation Coefficients, the second of our evaluation techniques, verified that the factors we measured correlated positively with each other. Clauses with few words in them tended to have few syntactic elaboration features, low gross weights and low embedded weights. And vice versa. (There were some surprises which we have noted but have not yet evaluated. For example, the incidence of syntactic elaborations that are not embeddings correlates with the incidence of conjunctions and with the incidence of embeddings themselves. But the incidence of conjunctions does not correlate with the incidence of embeddings. Why the first and not the second?)

In all, there were 612 orthographic units analyzed or 17 per composition (four compositions having been discarded for different reasons.) There were 806 clauses, or 1.30 clauses per sentence. The mean number of words per clause was 13.434, and the median was 11.847. There were 3.810 elaboration structures per clause, the median being 3.117. The mean and median figures for embeddings per clause were 1.026 and 0.590, respectively. The average weight score was 3.590, embedded weight accounting for 1.503 of this. The median weight score was 2.676, and the median embedded weight was 0.727.

Some clauses that the students wrote come very near these norms:

1. ...it is not possible to compare football with the brutality of gladiator fighting.
   (13 words; 4 elaborations: predicate adjective complement, prepositional phrase, prepositional phrase, noun noun; 1 embedding; gross weight, 3.25; embedded weight, 1.25.

2. ...the above paragraphs are an illustration of what I have experienced thus far in college.
   (15 words; 4 elaborations: prepositional phrase, 'what' nominalization,

---

*Just now, in fact, we are trying to straighten out difficulties in marking these last four. They are not easy to identify objectively, unambiguously.

adverb, prepositional phrase; 1 embedding; gross weight, 3.25; embedded weight, 1.50.

A sentence like "High school offers no challenge to brighter students" illustrates the unelaborated clause with zero score (also succinctness, a virtue which will go unrecognized until we have sophisticated the analysis considerably.) A sentence like the one below illustrates the syntactic earthquake (above 7.6 on the Jespersen scale.)

> I feel I must support Gleason's statement about a person giving himself away when he talks by tying in some aspects of his talk with perhaps his age, more than likely his education by choice of words and proficiency by which the talk is delivered and perhaps social background by maybe relating to some of his ways of life or lives of others that he knows.

The rhetorical analysis is presently in its final precomputer stage. The rhetorical features and logical relations have been marked, but the cards have not been punched. The only statistical values that can be reported today are those which we have estimated with paper and pencil, as follows. Analogy, polysyndeton, syntactic inversion, simile and apposition seem rare. We get one analogy and one asyndetic series every six compositions. There are 0.8 syntactic inversions per composition, one appositive, and 0.3 similes. Parallelisms are more common, about 1.6 of these per composition. But antithetical structures show up just 0.7 times. This same frequency is recorded for anaphoric expressions. Students use conscious metaphors 3.8 times, on the average, in their themes. Cliches (mercy me!) come in twos, twice blessed.

The variance in the incidence of logical connectors appears to be extremely wide. Some students use them abundantly (1.7 times per sentence!), others not at all (zero times per composition!) Perhaps it is here, more than anywhere else, that statistical averages will turn out to be the most misleading.

Students combine logical relations in apparently boundless permutations. They explain points previously made in whole sentences immediately following, and in series of sentences which, among themselves, express various other relations like additive, illustrative, adversative. They use illustratives as explanation, adversatives as alternatives, additives as illustrations. Their "logic," when it falls apart, does not break up <u>because</u> of the order of the elements necessarily, but because of their conduct of them. As yet, the program is capable of revealing little about what <u>standard</u> orderliness is. We shall see what it is perhaps, as we look more and more closely at how students arrange the parts of their argument and at the arrangements that professionals use.

In these remarks, I have wanted to persuade you that the task of measuring quantities of compositional structures appears to be both arduous and endless. (I am ready to accept volunteers for the job of analyzing a large sample of professional writing.) I have also wanted to suggest that the structures we have been counting and the quantitative comparisons we have thought to make may not be the ones which best reveal the key attributes of syntactic/rhetorical style or which decisively expose the differences between professional writing and freshman writing. I am not convinced that a quantitative representation of prose can divulge the secrets which it has kept from us. Nor am I convinced that the revealed secrets will help us to train writers. But I can tell you with enthusiasm that I would like to hear your views on these matters.

APPENDIX A

Weight	Code	Elaboration
0.25	AVD	Single Word Adverb
0.50	AJAJ	Adjective + Adjective
0.50	NON	Noun + Noun
0.50	POSV	Possessive Noun + Noun
0.75	ADJXX	Adjectival Phrase Type I
0.75	AVBXX	Adverbial Phrase Type I (time, place)
0.75	ELCNJ	Elliptical Conjunction
1.00	COMPAR	Comparative
1.00	NENN	Noun + Noun + Noun
1.00	PAPRT	Past Participial Phrase
1.00	PRSPT	Present Participial Phrase
1.25	ADJYY	Adjectival Phrase Type II (reduced relative clause, etc.)
1.25	AVBYY	Adverbial Phrase Type II (cause, direction, instrument, other)
1.25	PAC	Predicate Adjective Complement
1.25	PSS	Passive
1.25	VPC	Verb Phrase Complement
1.50	FOR	For-to Nominalization
1.50	ING	-ing Nominalization
1.50	NPHC	Noun Phrase Complement
1.50	REL	Relative Clause
1.50	THAT	That Nominalization
1.50	WH	What Nominalization
1.50	EMB2	Double Embedding
3.00	EMB3	Triple Embedding
6.00	EMB4	Quadruple Embedding
12.00	EMB5	Quintuple Embedding
1.50		Extra Weight For Nominalizations Used As Subjects

APPENDIX B

The Syntactic Structure of the Clause in Freshman Writing

	Number of words	Number of Embeddings	Number of Elaborations	Gross Weight	Embedded Weight
mean	13.434	1.026	3.810	3.590	1.503
median	11.847	0.590	3.117	2.676	0.727
range	66.000	10.000	26.000	27.750	15.000
variance	60.557	2.232	8.887	10.790	5.047